The trainability of abstract reasoning

van de Vijver, F.J.R.; Daal, M.; van Zonneveld, R.

Published in:
International Journal of Psychology

Publication date:
1986

Citation for published version (APA):
van de Vijver, F. J. R., Daal, M., & van Zonneveld, R. (1986). The trainability of abstract reasoning: A cross-cultural comparison. *International Journal of Psychology*, *21*, 589-615.

International Journal of Psychology 21 (1986) 589–615 589
North-Holland

THE TRAINABILITY OF FORMAL THINKING: A CROSS-CULTURAL COMPARISON *

Fons J.R. VAN DE VIJVER, Muriël DAAL and Renée VAN ZONNEVELD

Tilburg University, The Netherlands

Revised version received May 1986

In order to investigate what is learned in a training study on inductive reasoning, three tests were composed, involving two stimulus domains, namely letters and figures. Each test was composed on the basis of a set of explicitly defined item generating rules. The two figural tests were isomorphic tests, that is, they had identical problem structures but invoked different stimulus domains, simple geometric patterns and drawings of fruit. These tests were administered to pupils from grade 6 in The Netherlands, Surinam and Zambia. Half of each class was then trained at one of the three tests, while the other half did not get any training. During the training the pupils were taught to solve the items by trying to detect the features of the item generating rules. Four possible outcomes of transfer studies were postulated, differing in the generality of transfer of the results. By means of a Multivariate Analysis of Variance with Repeated Measures it was possible to differentiate between the four kinds of transfer; a pattern of fairly restricted transfer was demonstrated in each culture. Furthermore, in Zambia a remarkable score increase was found at the Letter Test, occurring both in the experimental and in the control condition. It was argued that this increase was caused by improved test-taking skills, learned at the first test administration.

There is a considerable amount of literature on the trainability of formal reasoning. Much of this work has been inspired by the observation that adolescents and adults often do not use formal operations in contexts in which this would be appropriate, e.g., in science instruction (cf. Chiappetta 1976). In virtually all investigations only Western samples were involved. The absence of cross-cultural training studies on formal reasoning is remarkable; after all, the problems of non-Western subjects to display formal structures in, e.g., testing situation has been amply documented (see e.g., Neimark 1975).

The main focus of most training studies is formed by the question

* Requests for reprints should be sent to F.J.R. van de Vijver, Dept. of Psychology, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

whether formal thinking in a particular domain belongs to the Vygotskian 'zone of proximal development', i.e., whether in a training individuals who apparently lack formal reasoning can be taught to use this kind of thinking. The empirical evidence on the trainability of formal reasoning is still inconclusive. Some investigators report substantial improvements in performance after a training (e.g., Kuhn and Angelev 1976; Siegler and Atlas 1976; Siegler and Liebert 1975). In other studies, however, far less favourable results were found (e.g., Bedderman 1973; Linn and Thier 1975). A reconciliation, though probably tentative, of this mutually contradictory evidence can be found in the work of Greenbowe et al. (1981). In an interesting replication of the studies by Siegler and his co-workers, these authors demonstrated that the training effects reported by Siegler c.s. were not durable over time and were test-specific, i.e., did not transfer to related, but untrained tasks. Therefore, these authors concluded that 'the training effect is better explained as application of memorized algorithms than as development of formal operations' (Greenbowe et al 1981: 705).

This latter study raises the important issue of what is being learned in a training of formal thinking, a question to which only cursory attention has been paid. Therefore, the present study has been designed in a way which allows for a detailed analysis of learning outcomes. To enable such an analysis, a taxonomy has been made of possible learning outcomes of a training. In the literature these are referred to as *transfer*. When test scores increase after a training, transfer is said to take place. Four different kinds of transfer will be distinguished, which vary along a dimension of psychological distance between the training and the learning outcome.

The first kind is called *similar transfer*, indicating a generalization from the test at which the subject is trained to parallel forms of the same test. The psychological distance between the training and the learning outcome is very small here; consequently, this will be the first kind of transfer which can be expected to occur. Gick and Holyoak (1983) call this 'reasoning from an analog', which is characterized by the fact that the subject's internal representation of the stimulus material will generally involve both particular stimulus features of the problems learned as well as more generalizable problem-solving strategies. Similar transfer will have a limited task generalizability as particular stimulus features are contained in it. It should be noted that similar transfer does not include mere practice effects, as transfer is a consequence of training, while practice effects, when present, will produce score increases in untrained individuals.

The second kind of transfer is called *near transfer* (cf. Royer 1979), which involves a generalization of the training results to other tests measuring the same psychological construct by means of problems, which are logically identical but involve different stimulus domains. In the literature such problems are called *isomorphic* (e.g., Kotovsky et al. 1985). As an example, suppose that an arithmetic test contains the item 'Two boys and two girls are playing. How many children are playing?'; an isomorphic item could be 'The family Doe has two sons and two daughters. How many children does the family Doe have?' When this kind of transfer occurs, the individual has learned a generalized problem-solving strategy, which is independent of the particular stimulus features used in the training, hence near transfer has a larger task generalizability than similar transfer.

The third kind of transfer, *far transfer* (cf. Royer 1979), has an even larger task generalizability. It is transfer to other tests measuring the same psychological construct in a different way. For instance, far transfer is said to occur when after a training in formal reasoning on Inhelder and Piaget's (1985) pendulum task, score increases are observed at other tasks for formal reasoning.

The fourth kind of transfer is *method transfer* or *test-wiseness*, which involves score increments to domains unrelated to the subject of the training. This can involve practice effects (Wing 1980), increased familiarity with testing, learning of particular response strategies, etc. (for a review see Sarnacki 1979).

The first three kinds of transfer are mutually related; the presence of far transfer presupposes the presence of near transfer, which in turn, presupposes the presence of similar transfer. In other words, broader kinds of transfer imply the occurrence of more restricted kinds of transfer. Method transfer at the other hand, is of a more general nature and can occur with any other transfer.

An analysis of learning outcomes will substantially profit from tests with explicit item gathering rules, so-called 'structured tests' (Van de Vijver in press). In the present study the tests administered have been constructed by means of Guttman's Facet Theory (e.g., Borg 1981; Cantor 1985). The basic idea behind Facet Theory is that 'variables included in a research design should be defined in terms of their component elements' (Foa 1965: 263). Instruments constructed by means of facet designs are called *structured tests*. As an example, suppose a researcher wants to compose an arithmetic test. First, a number of component elements are sought. The first component element of each item will consist of the kind of operation required, e.g., addition, subtraction, multiplication and division. In terms of Facet Theory this is called a facet with four levels. Suppose furthermore, that the researcher wants to use numbers with one, two and three digits in the test. This would constitute a second facet, with three levels. By forming the Cartesian product – called crossing in an Analysis of Variance – twelve kinds of items can be composed (four levels of the first facet × three levels of the second facet).

The use of facet designs offers two tightly linked advantages here. First, once a facet design has been composed, parallel tests can be easily made. In training studies both a pretest and a posttest have to be administered. These should be parallel tests. In Facet Theory this can be easily achieved by composing two tests with the same underlying facet design. Another advantage of Facet Theory is its allowance for constructing isomorphic tests, as these can be defined as tests with identical facet designs but dealing with different stimulus domains. Once a facet design has been composed isomorphic tests can also be readily made.

It should be emphasized that the outcome of a training may not be restricted to a single kind of transfer; rather, powerful training procedures will lead to more, i.e., more generalized kinds of transfer. However, in most studies only a subset of the kinds of transfer distinguished here are studied (e.g., the above mentioned studies). The present investigation has been designed in a way that allows for a simultaneous study of these four kinds of transfer. A detailed analysis will be made in the present study of what is being learned during a training of formal thinking. Furthermore, it is particularly interesting to compare these learning outcomes across cultures. The question raised then is whether subjects belonging to various cultural groups learn the same in a training.

Method

Samples

In three countries, The Netherlands, Surinam and Zambia, pupils from the sixth grade of primary school were involved. All the schools were situated in rural areas. Although the pupils from the same grade were involved in each country, the chronological ages are not identical. In The Netherlands and Surinam children start to attend the primary school at age six, while in Zambia children are at least seven years of age when they enter primary school.

The difference in cultural distance between these countries is considerable. Surinam, a former Dutch colony, and The Netherlands share a lot of common cultural heritage, not in the least in the educational system. Zambia on the other hand, is a Third World country with a large cultural distance to both Surinam and The Netherlands.

Tests

Three unspeeded tests for inductive reasoning were administered, one verbal and two figural tests. The former, called the Letter Test, is derived from the Letter Sets Test in the ETS-Kit of Factor-referenced Tests (Ekstrom et al. 1976). This test presumably reflects inductive reasoning in a symbolic medium. Each item in the Letter Test consists of five groups of six letters. Four out of these five have been generated according to some rule, while a fifth one does not follow this rule. The subject has to mark this latter group. In the Test consisting of 45 items the following five item generating rules (and only these!) are used:

(1) Each group of letters has the same number of vowels;
e.g., AEVZBR AIGVRS AVRTGK EUSRZQ UENBZR.
(2) Each group of letters has an equal number of identical letters which are the same across groups;
e.g., VVVVRG QZHGSL VTVVVB RFVVVV RFVVVV.
(3) Each group of letters has an equal number of identical letters which are not the same across groups;
e.g., RVGSBD CCCFHR WSWWGY CHKZZZ LMLLVS.
(4) Each group of letters has a number of letters which appear the same (*i.e.* 1, 2, 3 or 4) number of positions *after* each other in the alphabet;
e.g., BCDZRG QRSWVB KLMGSV FGHZDD VWHNKP.
(5) Each group of letters has a number of letters which appear the same (*i.e.* 1, 2, 3 or 4) number of positions *before* each other in the alphabet;
e.g., SVGTRP ZXVKBP NLJBWX FDBHVQ GRVTSZ.

In addition to this first facet, a second one was introduced, namely the numbers of letters to which the rule applies. This number varied from one to six. In the first example (about the vowels) the rule applied to two letters, while in the second example the rule refers to four letters.

The fourth and fifth rule refer to a difference in positions in the alphabet. These two rules imply an additional facet, dealing with differences of one, two, three and four positions in the alphabet. The example of rule four refers to a difference of one position and the example of rule five refers to a difference of two positions. The pretest version and its facet design, which is also the facet design of the posttest, appear in table 1.

The second test, called the Figure Test, is a figural completion test, presumably reflecting inductive reasoning in a symbolic medium. Each of the 30 items consists of five rows of twelve figures. The first eight figures are the same for all five rows, while the final

Table 1
The Letter Test and its facet design.

					Number of letters	Rule	Number of positions
1. GGGGGG	GGGGGG	GGGGGG	GGGGGP	GGGGGG	6	2	
2. XXPXXX	ZZRZZZ	CCCCCH	YDDDDD	FFFFPQ	5	3	
3. QIAEOU	WEIAVC	EIOUAZ	OUIEPA	WUEIOA	5	1	
4. HHHHHI	PPPPPP	RRRRRR	BBBBBB	JJJJJJ	6	3	
5. BCDEFQ	KLVMNO	RSUVWZ	FGHMIJ	TUVCWX	5	4	1
6. MLKJIH	GFEDCB	UTSRQP	ONMLKH	XWVUTS	6	5	1
7. BXDDDD	YDDDDD	DRDDDD	DDDSDD	DDYDDD	5	2	
8. FGHIJK	RSTUVW	ABCDFG	JKLMNO	PQRSTU	6	4	1
9. AAIOUU	IEEOOA	IIUUEA	AAUEEI	AEIOUZ	6	1	
10. PFEDCB	BWVUSR	KJIHGR	WVUQTS	JIHGLF	5	5	1
11. EAOIQU	WAOIEU	IOEXAU	AEPQIO	EAUVOI	5	1	
12. HHHHHI	HHHHHH	HHHHHH	HHHHHH	HHHHHH	6	2	
13. QQQQRQ	VSVVVV	TTTETT	BBABBB	DDYYDD	5	3	
14. AEEAIO	IIEESU	OOUEEI	IIAAEE	EEOOUI	6	1	
15. NNNNON	NNQNNN	RNNNNS	NNZNNN	YNNNNN	5	2	
16. RBBBBZ	XBCBBB	BBCBCD	RSBBBB	BBBBZZ	4	2	
17. GHHHHZ	KVKKKM	MMMPMT	RRRSTT	VVVVXY	4	3	
18. BCQDPZ	VWKXSZ	KLPPMM	QRLJVB	FGRHTY	3	4	1
19. VAEUSZ	EEOIPR	EIOUSV	UUEABG	OIEAQW	4	1	
20. PLRRWR	XVVVBD	NNYYTX	KKKQWR	DBDDFZ	3	3	
21. CXEGIK	RTKVXZ	KMOQBS	NPQRWZ	HJPLNP	5	4	2
22. OQTRPZ	GAEZRW	EDUXGK	IBWALV	GKAESW	2	1	
23. WMLRKP	DWVULG	SHGBFB	OUNMXK	BAFKRV	3	5	1
24. RRTURW	RRXSRA	WRXVRU	KRRBRK	RRDRPV	3	2	
25. LJHFDB	MKIGEC	ZXVTRP	PNLJHF	ECAACE	6	5	2
26. HIJTKZ	ACEGIK	PQBFRS	DJKLMV	RFSTUY	4	4	1
27. GIKMOQ	LNPRTV	PRTVXZ	ACEGIK	KMORVZ	6	4	2
28. FZIUEQ	RAIOXX	UORXDB	UECALW	PKEBIO	3	1	
29. DRPWQH	YBICJL	KSLBGS	PWNKRD	RAPBXZ	2	4	1
30. PNDLJH	SQOMKX	JHFEDB	ZXVTPR	STQMJG	5	5	2
31. RZBFJN	CGUZZG	HLPTDL	QNRVZB	EIMQBB	4	4	4
32. FAGHAR	ZVAWTA	AQMAGV	SAKVBR	DAMFAH	2	2	
33. QMIEAL	GERLYQ	WVSOKG	ZVRNMJ	TPLQHD	5	5	4
34. XROPVF	NVEWGN	HLPKVD	RCDIZT	PBGAWT	1	1	
35. DKLQWZ	KRDTEH	VHRNCX	FBHNSW	MBFOWK	2	4	2
36. KHEBVR	WTQNLL	JGDAWZ	UROLBK	XPLGFF	4	5	3
37. WBMLVR	DTFHSY	DCQNLR	HRLZSD	YBLWSP	1	2	
38. XDBKRE	TNLLRK	VEGVSB	PFCCHL	OZAGZQ	2	3	
39. KVRSWU	PSDVYB	DGJTMN	FIXLRO	QTPWDZ	4	4	3
40. TRNJFV	XTPLDS	MIEAQT	VKOKGC	PFDVXB	4	5	4
41. PVXFDB	SVRYYD	QWZLJH	GHPANL	RZWWXV	3	5	2
42. VRNJFB	ZTMGDB	XTPLHD	UQMIEA	WSOKGC	6	5	4
43. AYXKWV	BNMLPK	BCDFGH	IHGFZR	YUWTSR	4	5	1
44. FPNMXY	UDMCFP	PSRDKN	KVLLWC	GYXVJB	2	5	1
45. YVPHDA	CRBFOO	WGKGSV	KLRVJQ	OSFDPZ	2	4	4

four differ in each row. One of the five rows has been entirely formed according to some rule. The subject has to mark this latter row. Each figure possesses a minimum of one and a maximum of six elements. A figure consists of either a square or a circle; furthermore, a hat, an arrow, a dot, a dash and a bow can form part of the figure. The position of these elements in the figure is always the same (see fig. 1 for an illustration).

The following facets are included in the test:

(1) The number of figures in a period, i.e., the number of circles or squares following each other, can be two, three or four. (Periods are formed by concatenating either subsequent circles or circles, but a period never contains both.) In the first item of fig. 1 there are four figures and in the second item there are three figures in a period.
(2) The number of elements in which subsequent figures in a period differ, can be one, two or three. In the first item of fig. 1 subsequent figures in a period differ in one element and in the second item in two elements.
(3) Three different item generating rules (and only these) have been used:
   - One or more elements are added to subsequent figures in a period (cf. the first example of fig. 1).
   - One or more elements are subtracted from subsequent figures in a period.
   - Alternatingly, one or more elements are added to and subtracted from subsequent figures in a period (cf. the second example of fig. 1).
(4) The variation across periods, i.e., the elements added or subtracted, may be the same or differ. In each item of fig. 1 there is a constant variation.
(5) Subsequent periods of an item may or may not repeat each other, i.e., the first figure of each group are the same except for circles and squares which are interchanged. In the first example there is a repetition, while this is not the case in the second example. The latter two facets are not independent; a repetition of groups is only possible with a constant variation of the fourth facet.

The facet design matrix of the test is given in table 2. Because it would take too many items to cover each cell of the design matrix, a selection of 30 items was made. The five facets mentioned above amount to a design matrix of 13 columns, three columns referring to the first facet (two, three or four figures in a period, respectively), three columns referring to the second facet, three columns for the item generating rules, and two columns for the final two facets. This facet design matrix has been used to generate the items of both the pretest and the posttest, analogously to the procedure for the Letter Test.

The third test is the Fruit Test, which is an isomorphic version of the Figure Test. This means that the facet design matrix of table 2 has been used to generate the items of the pre- and the posttests of the Figure Test and the Fruit Test. The most important difference between the Figure and the Fruit Test consists of the concreteness and probably the familiarity of the stimulus materials. In the former abstract elements are used in the figures (circles, squares, lines, etc.). In the Fruit Test these abstract elements were replaced by drawings of dishes with fruits on them. There was a one to one correspondence between each figure of the Figure Test, both of the pre- and the posttest, and the Fruit Test. This was achieved by using a simple set of (arbitrary) transformation rules between the elements of the two tests. An example of an item from this test is given in fig. 1. Note that the top and the bottom item of fig. 1 are isomorphic items of the Figure and Fruit Test, respectively.

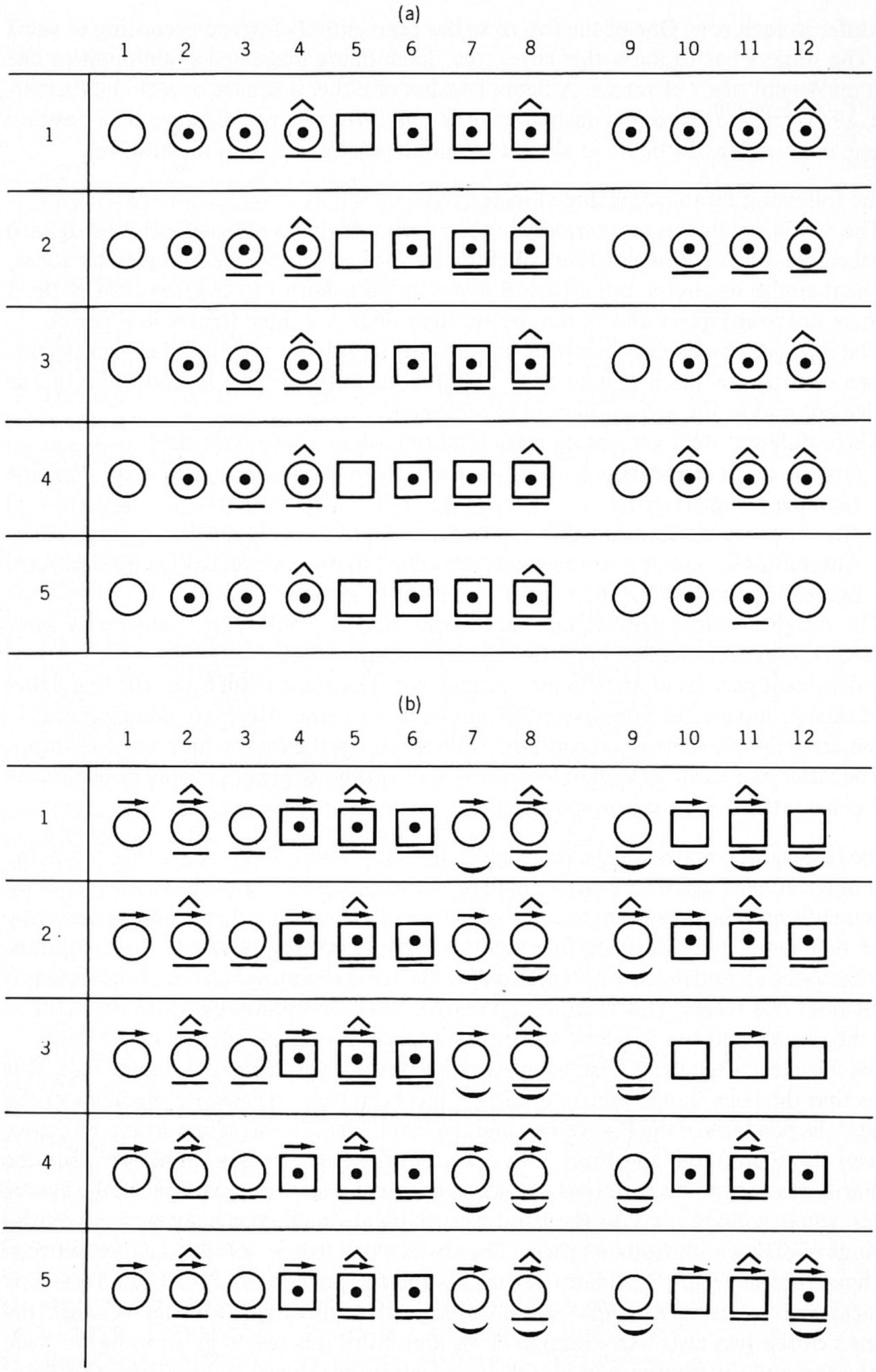

Fig. 1. (a, b)

(c)

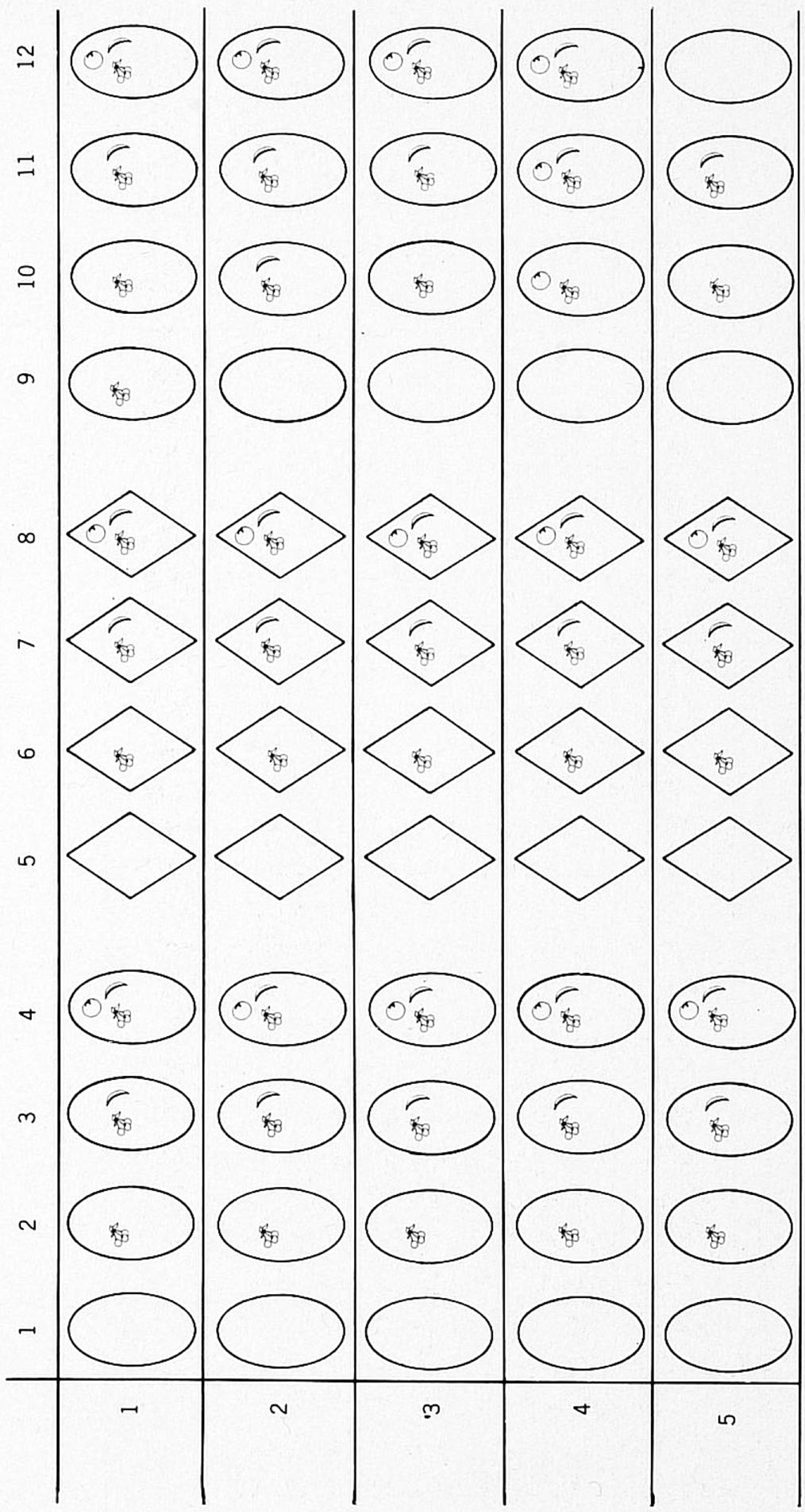

Fig. 1. Some examples of items of the Figure Test (a, b) and the Fruit Test (c).
Note: a and c are isomorphic items.

Table 2
The facet design of the figural tests.

	Period	Variation	1 aspect varied			2 aspects varied			3 aspects varied		
			Rule 1	Rule 2	Rule 3	Rule 1	Rule 2	Rule 3	Rule 1	Rule 2	Rule 3
Groups repeat each other	2	Constant	1	2		5	9		6	14	
		Variable									
	3	Constant	8		15	7	12	11			
		Variable									
	4	Constant	3	4	13						10
		Variable									
Groups do not repeat each other	2	Constant								19	
		Variable		24		30		26			
	3	Constant		22			17				18
		Variable	21					25			
	4	Constant	16	29				20			
		Variable			23						27

Design

These tests for inductive reasoning were administered in a pretest – training versus control group – posttest design, with each involved class randomly split up into two groups. Half of each class got a training, while the other half did not get any training. Composing the experimental and control groups in this way has the advantage that class differences are not confounded with differences between control and experimental groups, which occur when the same treatment is given to classes as a whole rather than to random splits of a class.

Three experimental conditions were implemented, which differed in the subject of training. For each of the three kinds of tests there was a corresponding experimental condition. Subjects in the first experimental condition got a training at the Figure Test; other subjects were training at the Letter Test, while a third group received a training at the Fruit Test.

To each subject, both in the experimental and the control condition, the three tests for inductive reasoning were administered at both the pre- and the posttest. The Dutch and Surinam pupils were tested by the second author; the Zambian subjects were tested by the third author. In order to control for serial effects, the tests were administered in a random order on the pretest and in the counterbalanced order at the posttest. In the three countries the same test order was maintained. The design of the study, including the sample size per condition and per country, are schematically given in table 3.

The administration of each test at the pretest sessions was preceded by a lengthy introduction in which a demonstration was given of each of the item generating rules. After that, the pupils were asked to answer exercise items, which, again, covered all the rules. The tester explained the correct answer. After this instruction the subjects were asked to answer the actual test items. This lengthy instruction, with an explicit statement of the valid item generating rules, was given in order to avoid any confusion about the problem space of the test (cf. Van de Vijver in press).

The testing and training was done at three consecutive school days. The first version of each test was administered at the first day. At the second day a group of pupils, who

Table 3
The design of the study.

Training at	Pretest	Posttest	Sample size		
			Dut	Sur	Zam
Figure Test	Le, Fr, Fi	Fi, Fr, Le	13	15	17
Control Group	"	"	15	16	13
Fruit Test	Fr, Fi, Le	Le, Fi, Fr	14	16	18
Control Group	"	"	11	12	16
Letter Test	Fi, Le, Fr	Fr, Le, Fi	15	15	18
Control Group	"	"	9	15	17

Note: Fi = Figure Test; Fr = Fruit Test; Le = Letter Test.

were randomly selected by the experimenter, were trained; posttest assessment took place at the third day.

The training of the experimental subjects took one session. This training included a thorough treatment of and exercises in not yet administered items of one of the kinds of tests (Letter, Figure or Fruit) depending on the experimental condition. In this training the pupils were taught to solve the items in a formalized way by detecting the rule of the item; that is the subjects were trained in a strategy of searching formal properties of items. Analogously, in the training at the Letter Test, subjects were taught to use the strategy of disentangling each item into the facets of table 1: which rule applies? and to how many letters does this rule apply in this item? Analogously, subjects, who were trained at the Figure or the Fruit Test, were instructed to solve the items by answering the questions: how many figures are there in a group? which rule does apply? and how many elements are added or subtracted? In other words, in the training at each test the subjects were instructed to look for the item generating rule behind the item. The training sessions took approximately half a school day.

Operationalization of transfer

On the basis of the previously described design operational definitions can be given of the four different kinds of transfer distinguished in the introduction. *Similar transfer* is assumed to occur when after a training an increased test score is observed at the trained test, which is not also observed to the same amount in the control condition at the same test. *Near transfer* occurs when after a training at the Figure or Fruit Test an increased test score is observed at the Fruit or Figure Test respectively, which is not also observable to the same amount in the control condition. *Far transfer* is present when after a training at the Letter Test or the Figure/Fruit Test an increased test score is observed at the Figure/Fruit Test or the Letter Test respectively, which is not also found to the same amount in the control condition. Finally, *method transfer* is said to take place when in the control group a score increase can be observed at one or more posttests.

Results

From a theoretical point of view it would be expected that the three tests are mutally interrelated, as each of them presumably measures inductive reasoning. This claim was investigated in an analysis from which the results are given in table 4. Two kinds of relations are reported: the correlations in the concatenated samples and the averaged correlations of the three within-culture correlation matrices. Both types of correlations are reported as they need not be the same; particularly when large intercultural differences in test scores are observed, the correlations in the concatenated samples can be seriously influenced. As can be seen in table 4, both correlation matrices contain acceptably high values, which substantiates the theoretical expectation of a unidimensional trait underlying the test performance at the various tests, although there are rather large differences in the correlations below and above the diagonal of table 4. As will be seen later, these differences are caused by large intergroup score differences.

In an additional first check of the data, the reliability of the tests was computed for

Table 4
Intertest correlations.

	(1)	(2)	(3)	(4)	(5)	(6)
(1) Figure Test pre	–	0.36	0.40	0.59	0.62	0.51
(2) Figure Test post	0.71	–	0.52	0.38	0.30	0.40
(3) Fruit Test pre	0.81	0.71	–	0.50	0.33	0.48
(4) Fruit Test post	0.76	0.84	0.75	–	0.56	0.76
(5) Letter Test pre	0.67	0.64	0.61	0.60	–	0.50
(6) Letter Test post	0.59	0.62	0.54	0.62	0.64	–

Note: The values below the diagonal are the correlations for the concatenated samples; the values above the diagonal are averages of the within culture correlations.

Table 5
Average scores at the tests in each culture.

Training at	Exp./ control	Culture	Test					
			Figure		Fruit		Letter	
			Pre	Post	Pre	Post	Pre	Post
Figure	Exp	Dut	14.07	19.00	13.33	15.33	27.47	23.73
		Sur	10.13	14.40	10.60	14.40	23.53	22.47
		Zam	5.00	7.76	4.65	7.47	9.47	17.82
	Cont	Dut	17.62	18.38	14.92	15.15	30.77	28.38
		Sur	9.25	12.44	10.00	10.63	24.13	21.56
		Zam	4.54	5.92	4.53	6.23	6.46	12.62
Fruit	Exp	Dut	15.43	19.43	14.14	16.21	24.71	28.07
		Sur	11.19	14.06	10.81	12.44	14.13	20.56
		Zam	8.39	9.72	6.44	9.17	8.94	18.94
	Cont	Dut	13.82	15.82	10.82	13.00	20.00	25.45
		Sur	9.00	10.67	8.25	9.08	15.58	19.92
		Zam	4.13	5.75	3.69	6.69	6.50	15.69
Letter	Exp	Dut	19.93	17.40	18.60	21.07	21.87	28.73
		Sur	9.47	10.07	9.27	10.07	15.53	20.00
		Zam	8.89	11.78	8.72	11.17	15.50	25.22
	Cont	Dut	20.89	18.22	16.89	21.78	23.00	25.22
		Sur	11.27	11.07	8.67	11.13	17.13	19.40
		Zam	5.24	10.06	5.53	9.88	12.00	21.35

each cultural group. The KR-20 turned out to be satisfactory; they ranged from 0.55 to 0.90, with an average value of 0.76. The reliability of the three tests was approximately the same, even though the Letter Test has 15 more items. This confirms earlier findings with these tests (Van de Vijver in press). It was also found that the reliabilities of the tests were somewhat higher in The Netherlands, due to the fact that the dispersion in the test scores was higher in this sample.

As a final check of the data, the score equivalence of the tests was investigated by means of a procedure described by Van de Vijver and Poortinga (1982). For each test a dichotomous data matrix (Cultures by Subjects by Items) is composed, in which the contribution of the Item by Culture interaction, the presumed bias indicator is evaluated by means of its unit sampling generalizability coefficient. The average value of this coefficient turned out to be 0.03, which indicates that item bias has not substantially influenced the data.

On the basis of the three checks on the adequacy of the present data set, consisting of intertest correlations, test reliabilities and a bias statistic, it can safely be concluded that no major anomalies have occurred and that the data can be meaningfully analyzed with respect to the presence of transfer.

In table 5 the average test scores are presented which were obtained in the various groups in each culture. The cross-cultural differences in these scores appeared to be large, which is very common in this kind of research (Rogoff 1981). As in our design no explicit provisions were made to relate such score differences to background characteristics of the subjects (cf. Van de Vijver and Poortinga in press), no attempt will be made to elaborate upon them.

One of the major questions of the present study involves the problem what kind of transfer is observed in a training of inductive reasoning. In order to get insight into this problem, the data of table 5 have been analyzed by means of a Multivariate Analysis of Variance with Repeated Measures (the program P4V in the BMDP-series, Dixon 1981). For each of the three training conditions a separate analysis was carried out. The independent variables in these analyses were Culture (Dutch, Surinam, and Zambian) and Training (Experimental and Control) as the Between-Subjects Factors and Session (Pretest and Posttest) as the Within-Subjects Factor; the test scores at the three pre- and posttests were the dependent variables. The results of the analyses are given in table 6, in which for the clarity of the presentation, only the significance levels are presented.

Similar transfer, previously defined as a score increase after training at the trained test, which is not observed in the control group as well, will come out in the present analyses as a significant Session by Training interaction. It can be seen in table 6 that this interaction was significant for the Figure Test ($p < 0.04$) and the Letter Test ($p < 0.05$), but not for the Fruit Test ($p = 0.89$). An inspection of table 5 shows that training at the Fruit Test also gave rise to an increase of the mean score in each sample (2.07 in the Dutch, 1.63 in the Surinam, and 2.73 in the Zambian sample). However, in the control condition the size of the increase was comparable (2.18, 0.83, and 3.00, respectively), hence no significant Session by Training interaction could be observed. Furthermore, it appears from table 6 that training at the Fruit Test was the only experimental condition with a significant difference between experimental and control subjects; at both test sessions the experimental group scored higher than the control group at each test. Thus, it seems that

Table 6
Results of the multivariate analyses of variance with repeated measures.

Experimental factor	Experimental condition		
	Figures	Fruits	Letters
Culture (C)			
General	0.00	0.00	0.00
Figure Test	0.00	0.00	0.00
Fruit Test	0.00	0.00	0.00
Letter Test	0.00	0.00	0.00
Training (T)			
General	0.67	0.01	0.21
Figure Test	0.71	0.00	0.89
Fruit Test	0.32	0.00	0.36
Letter Test	0.92	0.04	0.12
$C \times T$			
General	0.03	0.21	0.35
Figure Test	0.42	0.73	0.11
Fruit Test	0.32	0.95	0.48
Letter Test	0.00	0.22	0.15
Session (S)			
General	0.00	0.00	0.00
Figure Test	0.00	0.00	0.26
Fruit Test	0.00	0.00	0.00
Letter Test	0.14	0.00	0.00
$S \times C$			
General	0.00	0.00	0.00
Figure Test	0.42	0.53	0.00
Fruit Test	0.54	0.35	0.12
Letter Test	0.00	0.00	0.00
$S \times T$			
General	0.05	0.82	0.05
Figure Test	0.04	0.38	0.69
Fruit Test	0.02	0.89	0.02
Letter Test	0.46	0.81	0.05
$S \times C \times T$			
General	0.42	0.74	0.67
Figure Test	0.42	0.68	0.35
Fruit Test	0.61	0.88	0.94
Letter Test	0.37	0.34	0.37

Note: The numbers in the cells represent significance levels.

the lack to find a significant Session by Training interaction seems to be caused to a larger extent by sampling fluctuations than by the non-occurrence of transfer.

The second kind of transfer distinguished is near transfer, which refers to score increases at the Figure and Fruit Test after a training in the Fruit and Figure Test, respectively. In the present analyses this kind of transfer will also lead to a significant Session by Training interaction, as was postulated for similar transfer. For near transfer, however, the score increase should be observed at a different dependent variable; when near transfer occurs, training at the Figure/Fruit Test leads to a significant interaction at the Fruit/Figure Test, respectively. From table 6 it can be seen that for the Figure Test the Session by Training interaction is significant ($p < 0.02$). For the training at the Fruit Test, however, the problem of the apparent group differences between the experimental and the control subjects also arises here. An inspection of the mean score increases at the Figure Test after training at the Fruit Test (see table 5) yields a picture which seems to favour the interpretation of near transfer: 4.00 in the Dutch, 2.87 in the Surinam, and 1.33 in the Zambian sample. These values are slightly higher than the increases in the control groups: 2.00, 1.67, and 1.62, respectively.

The third kind of transfer, far transfer, was defined as score increases at the Letter Test, c.q. figural tests after a training at the Letter Test, c.q. one of the figural tests. This kind of transfer will also manifest itself as an interaction between Session and Training: for instance, a training at the Figure Test will have induced far transfer when this interaction is significant in an analysis with the Letter Test scores at the pre- and the posttest as the dependent variables. From table 6 it appears that training at the Figure Test does not lead to far transfer ($p = 0.46$); training at the Letter Test does lead to significant increase at the Letter Test ($p < 0.02$), but this is not the case for the Figure Test ($p = 0.69$). With respect to the Fruit Test, the same reservations should be kept in mind as has been done previously. But even with this ambiguity in mind, there does not seem to be strong evidence for the presence of far transfer.

The fourth kind of transfer, method transfer, is present when in the control group the average scores increase from the pretest to the posttest. In order to investigate this, another Multivariate Analysis of Variance with Repeated Measures was carried out; Culture (Dutch, Surinam and Zambian) and the Training Condition (Figures, Fruits, Letters) with which the particular control group was matched, were the Between-Subjects factors in this analysis and Session (pre and post) was the Within-Subjects factor. The test scores at the first and second administration of the three tests were the dependent variables. In this analysis method transfer will reveal itself in the main effect for Session, when the transfer affects the performance of the subjects in all cultures, and in the interaction between the subjects in all cultures, and in the interaction between Session and Culture, when the transfer only affects one or two cultures. The main effect for Session turned out to be highly significant for each test ($p < 0.0001$), indicating the effects of method transfer at each test in the three cultural groups. It was also observed that the Session by Culture interaction was highly significant ($p < 0.0001$) for the Letter Test. This will be discussed later on.

From the analyses thus far, it appears that similar transfer, near transfer and method transfer have probably affected the data. All these kinds of transfer result in score increases from the pretest to the posttest. Apparently, transfer affected the scores of both the experimental and the control subjects, thereby inducing overall score differences between the pretest and the posttest. When these increases are sufficiently large, the main effect for Session will become significant. In fact, most of the tests of the main effect for Session turned out to be significant (see table 6). Thus, it seems that the main effect for Session is a contamination of different kinds of transfer.

It could be argued that this interpretation of the outcomes of the Analysis of Variance is debatable; after all, the use of this technique in the measurement of change is not without problems, in particular when the pretest performance levels between the groups are considerable, which was clearly the case in the present data set. One of the problems which can occur is that the experimental effects are overestimated. As a consequence, the choice for the Analysis of Variance model may seem less than optimal; a latent trait approach, for instance the application of the Linear Logistic Test Model as done by Van de Vijver (in press) would be a better choice. In our opinion, this argument would hold if we would not extend the analyses beyond the ones already presented. In subsequent analyses, however, the validity of the interpretations of the Multivariate Analysis of Variance will be carefully scrutinized.

In these subsequent analyses a closer look is taken at the transfer which has affected the data. The facet designs implemented in the tests offer a good opportunity to analyze transfer at a meaningful subtest level. In the analyses to be reported the major variable is the average score at a particular facet level. This variable, denoted by p_k, is defined as:

$$p_k = \sum_i \sum_j u_{ijk} / N \times n_k,$$

in which $u_{ijk} = 0$ if subject i answered them j, which belongs to facet level k, incorrectly and $u_{ijk} = 1$ if the subject answered the item correctly; N is the sample size and n_k is the number of items pertaining to facet level k. For instance, in computing the average score for the facet level 'vowels' of the Letter Test, the (dichotomously scored) responses of all subjects on all items about vowels (this can be one up to six vowels) are summed and then divided by six (different kinds of items) $\times$ N (being the sample size). The variable p_k has a straightforward interpretation; it is the empirical probability of obtaining a correct response at facet level k, in the example the probability that in a particular group an item about vowels is answered correctly. In other words, p_k is the p-value of facet level k. As training effects are the primary focus of the present study, the increase in p_k-values from pre- to posttest will be reported rather than the p_k-values as such. The increase in p_k-values for the three tests will be reported for the experimental group and the control group, separately. In order to increase the stability of the results, scores of all control subjects from the three classes rather than from one class have been taken together.

The increase in p_k-values at the Figure Test are reported in table 7. It can be clearly seen in this table, that in each culture the increases are larger in the experimental groups than in the control groups. This replicates the observation of a significant Session by Training interaction in the previously discussed Multivariate Analysis of Variance. It can also be observed in the table that the mutual differences between the facet levels within an experimental group are rather small; this means that all the relevant facets of the test are mastered more adequately after the training. Spada and McGaw (1985) have called this training result, in which subjects increase their knowledge of all the relevant concepts of the test 'global learning'. It also appears from table 6 that this pattern of 'global learning' can be found in each cultural group, even though the increase in average

Table 7
Score increases for each facet level at the Figure Test.

Facet	Culture					
	Dutch		Surinam		Zambian	
	Exp	Con	Exp	Con	Exp	Con
Periodicity cues						
Present	0.16	0.03	0.14	0.03	0.10	0.10
Absent	0.17	−0.03	0.15	0.09	0.07	0.08
Number of figures in a period						
2 figures	0.11	−0.05	0.15	0.01	0.14	0.12
3 figures	0.23	0.03	0.14	0.04	−0.01	0.07
4 figures	0.15	0.05	0.14	0.11	0.17	0.09
Variation across periods						
Constant	0.19	0.04	0.15	0.07	0.10	0.09
Variable	0.10	−0.06	0.12	0.02	0.07	0.10
Rule which applies						
Rule 1	0.15	0.01	0.14	0.05	0.13	0.11
Rule 2	0.13	−0.02	0.11	0.04	0.07	0.12
Rule 3	0.20	0.04	0.19	0.07	0.10	0.06
Number of aspects varied						
1 aspect	0.16	0.01	0.09	0.04	0.12	0.09
2 aspects	0.26	0.05	0.13	0.11	0.04	0.11
3 aspects	0.11	−0.02	0.23	0.02	0.11	0.06

p_k-values varies considerably across cultures. This means that there are no major cross-cultural differences in learning outcome of the training at the Figure Test. A final relevant aspect of table 7 concerns the increase in p_k-values for the control groups. As most entries of the control condition are positive, in particular for the Surinam and Zambian sample, it can be concluded that method transfer affects the data, as was observed earlier. From the present analysis it can also be derived that global learning also takes place in this condition.

In table 8 the same procedure has been followed for the Fruit Test. From an inspection of this table it is clear that the results of the Figure Test are largely replicated, which is not surprising for two isomorphic tests. Again, global learning seems to take place. For the Fruit Test the pattern of a general increase in average p_k-values is even more pronounced; there is only one negative entry in the table!

The gain in average p_k-values at the Letter Test is presented in table 9. The most striking finding is the tremendous increase of the average p_k-values in the Zambian sample. In other words, both the trained and the untrained Zambian subjects profit considerably from earlier experiences with the test; in the Multivariate Analysis of

Table 8
Score increases for each facet level at the Fruit Test.

Facet	Culture					
	Dutch		Surinam		Zambian	
	Exp	Con	Exp	Con	Exp	Con
Periodicity cues						
Present	0.06	0.06	0.07	0.07	0.07	0.12
Absent	0.09	0.10	0.03	−0.00	0.13	0.08
Number of figures in a period						
2 figures	0.03	0.11	0.02	0.03	0.10	0.12
3 figures	0.09	0.08	0.10	0.07	0.08	0.11
4 figures	0.10	0.01	0.03	0.03	0.10	0.08
Variation across periods						
Constant	0.05	0.07	0.06	0.05	0.07	0.12
Variable	0.12	0.09	0.05	0.03	0.15	0.07
Rule which applies						
Rule 1	0.03	0.10	0.07	0.03	0.11	0.13
Rule 2	0.11	0.04	0.03	0.05	0.00	0.11
Rule 3	0.07	0.06	0.06	0.06	0.17	0.07
Number of aspects varied						
1 aspect	0.08	0.05	0.04	0.01	0.07	0.10
2 aspects	0.08	0.05	0.02	0.08	0.10	0.10
3 aspects	0.03	0.14	0.13	0.07	0.13	0.11

Variance of the Letter Test this came out as a highly significant Session by Culture interaction. In each cultural group the entries of table 9 of the experimental groups tend to be larger than of the control groups, which has also been found at the figural tests. This indicates that for each experimental condition similar transfer accounts for larger score increases than method transfer does. Another aspect of the table worth mentioning concerns the differences of the entries between facet levels within a single facet. These differences are more pronounced in this test than in the previous ones. It appears, for instance, that experimental subjects in the Dutch and Zambian sample and to a somewhat lesser extent in the Surinam sample learn to solve items about vowels (rule 1) with increasing success; the probability that these subjects answer such an item adequately is 0.30 larger at the posttest in the former two samples than at the pretest. This is the result of the learning of the application of the concept 'vowel'. Apparently, the global learning pattern does not apply here; experimental subjects learn to apply particular item generating rules more successfully. As shown in the Zambian control group, mere practice can induce the same result of specific learning.

The results with respect to the final facet of table 9, the number of positions in the alphabet, reveals an interesting picture. In the Dutch sample the facet levels referring to

Table 9
Score increases for each facet level at the Letter Test.

Facet	Culture					
	Dutch		Surinam		Zambian	
	Exp	Con	Exp	Con	Exp	Con
Number of letters						
1 letter	0.20	0.08	−0.17	0.05	0.17	0.16
2 letters	0.10	0.06	0.22	0.03	0.19	0.14
3 letters	0.17	0.07	0.10	0.05	0.38	0.27
4 letters	0.20	0.01	0.04	−0.04	0.24	0.18
5 letters	0.21	−0.01	0.17	0.05	0.18	0.19
6 letters	0.12	0.05	0.07	0.02	0.16	0.16
Rule which applies						
Rule 1	0.30	0.09	0.12	0.09	0.34	0.22
Rule 2	0.07	0.04	0.08	−0.03	0.17	0.29
Rule 3	0.14	0.01	0.24	0.06	0.16	0.25
Rule 4	0.05	0.06	0.06	0.04	0.28	0.12
Rule 5	0.21	−0.02	0.07	−0.02	0.13	0.12
Number of positions (rule 4 and rule 5)						
1 position	0.13	0.01	0.16	0.04	0.30	0.14
2 positions	0.10	0.04	0.06	0.01	0.21	0.15
3 positions	0.20	−0.03	−0.07	−0.03	0.06	0.01
4 positions	0.16	0.01	−0.05	−0.02	0.08	0.08

the higher number of positions seems to benefit most from a training. These facet levels represent the most difficult items of the test. In the Surinam and the Zambian sample, however, the facet levels referring to less letters profit the most. The overall picture for the three samples can be summarized by stating that samples with relatively low scores at the pretest master the easier facet levels, while in the score groups with a higher initial score level there is not very much room for improvement at the easier facet levels; this latter group of subjects gradually master the more difficult facet levels, while these levels remain beyond the reach of the lower scoring groups.

Discussion

In the present study evidence was found that a training of abstract reasoning induces similar transfer at the two figural tests and the Letter Test, near transfer at the two figural tasks, which consisted of isomorphic items, and method transfer at the two figural tests and the the Letter Test. This pattern was found in each cultural group, even though the size of the transfer differed across groups; for instance, method transfer was most pronounced at the Letter Test in the Zambian group. In none of the groups strong evidence could be found for the presence of far transfer.

From a psychological point of view these findings mean that the training induced a complex set of item solving strategies, some of them being restricted to the stimulus domain of the trained test, called 'reasoning from an analog' by Gick and Holyoak (1983), while other strategies are of a more general nature, called 'reasoning from a schema' by these authors. When similar transfer ('reasoning from an analog') occurs, the strategies developed will probably involve both general aspects, e.g., an increased skill to look for the three item generating rules at the figural tests, and test-specific features, e.g., an increased skill to decode subsequent figures of a figural test into different periods. In near transfer ('reasoning from a schema') only the general aspects are involved. It should be emphasized, however, that, as stated in the introduction, the different kinds of transfer differ in degree rather than quality. The stimulus boundedness of the strategies mastered will vary from one individual to another and the labeling of these strategies as a particular kind of transfer is only meant as a summary of a complex psychological process.

The present findings are in line with earlier investigations (e.g., Greenbowe et al. 1981; Holyoak 1985; Wing 1980). A similar close link between item generating rules, test design and training as in the present study can be found in the work of Holzman et al. (1976). On the basis of Simon and Kotovsky's empirical work and computer simulations on serial pattern learning (Simon and Kotovsky 1963; Kotovsky and Simon 1973), Holzman et al. trained children on tests for serial letter patterns by learning them to detect the relevant item features, much in the same way as has been done in the present study. These authors also found substantial score increments when the children had mastered a strategy of looking for the relevant item generating rules.

Some evidence was found for the differential transfer across the tasks of the present study. The figural tests yielded a picture of global learning, while at the Letter Test more item-specific learning seemed to take place. In our opinion, this finding is a consequence of the nature of the test facets and is not inherent to the stimulus domains. Although not further documented in the previous section, it was found in the analyses that within a single facet the dispersion of the facet levels differed considerably across the tasks. More specifically, at both figural tests this dispersion was lower than at the Letter Test. In other words, the differences between the easiest and most difficult level of the facets were larger at the Letter Test than at the figural tests. Moreover, the overall dispersion of the average facet difficulties (i.e., taken over all facets simultaneously) was smaller at the figural tests. It is not surprising that on a test with relatively homogeneous facet level difficulties score increases influence all facet levels in more or less the same degree, as found here. Global learning is more likely to occur at tests with such homogeneous facet level difficulties, while more specific learning is more likely to occur at tests with heterogeneous facet level difficulties, as found for the Letter Test. The impact of the average item difficulty of the pre- and posttest at the outcome of the training study is probably somewhat unexpected, even though not negligible. In addition to the heterogeneity of the item difficulties, floor and ceiling effects can seriously influence the degree of transfer observed. An important consequence of the present findings is that the content of what is being learned in a training study can vary considerably from one task to another, even when the tasks measure essentially the same construct, inductive reasoning in the present case.

The observation of relatively homogeneous facet level difficulties also explains why the test reliabilities were higher for the figural tests than for the Letter Test. In tests with homogeneous facet level difficulties subjects will tend to be able to answer all the questions correctly or incorrectly. As a consequence, the dispersion of the test scores will be high in comparison with a test with heterogeneous difficulties and hence, the reliabilities of the figural tests will be higher.

It was found with a remarkable consistency in this study that the performance differences between the figural tests were minimal. This was observed in each cultural group under study. Furthermore, a similar pattern of transfer could be observed for both tests. This means that the major difference between these tests, i.e., the concreteness and probably the familiarity of the stimulus material, does not affect the performance of the subjects to a large degree. This seems counterintuitive at first glance, as it has been found over and again in cross-cultural psychology that stimulus familiarity can influence the test performance dramatically (e.g., Van de Vijver and Poortinga in press). However, in this study the most concrete stimulus material, drawings of fruit, was not used in its natural context as would be the case, for instance, in a classification task or any other task in which knowledge of the stimulus could be put to an advantage. This did not hold for the present tests. In an item any set of the fruits could go together and make up a meaningful set, while the underlying rules were similar for the familiar and the unfamiliar task. So, subjects could not benefit from their acquaintance with the stimuli. A potentially important principle for cross-cultural test design can be derived from this observation: stimulus familiarity is only important to the extent that subjects can benefit from the knowledge based on it. An analogous argument holds for the inclusion of 'culture free' stimulus material in a test; increasing the unfamiliarity of stimulus features, e.g., as in Cattell's Culture Fair Intelligence Test, will not diminish cultural differences.

Another aspect worth mentioning with regard to the present study concerns the high degree of method transfer in the Zambian group. Such a finding is an implicit threat of the comparability of test results (Poortinga 1975), even though the item bias was found to be low. The present data do not meet the requirements of score comparability when this is taken to mean that test scores have the same psychological meaning. The test performance of the Zambian subjects is apparently influenced by other, additional factors which affect the test performance of the Dutch and Surinam subjects to a considerably lower degree. On the basis of the presence of method transfer it can be argued that the Zambian subjects learned a number of the skills needed in the tests at the pretest. In comparisons between Western and non-Western groups there will be considerable cross-cultural differences in the degree of mastery of the test-taking skills. However, within a single cultural group individuals will have a rather homogeneous standing on them. Once again, the potential impact of test-wiseness on test behaviour is demonstrated (cf. Rogoff 1981).

In a detailed study of which facet levels profited most from the training and/or retesting, cross-cultural differences were found. It is essential to emphasize here that these cross-cultural differences are mainly a function of the performance level at the pretest and hence, that transfer is a function of score level rather than culture. In our opinion it is likely that a prolonged training of the Zambian and Surinam samples at the Letter Test would result in a pattern of score increase, which was observed in the Dutch sample.

Some reservation should be made about the replicability of the results of the present study. First, it should be kept in mind that our training procedure was short and hence, it would be unrealistic to expect a very broad transfer after it; only after a more prolonged training far transfer can be expected. Another reservation about the present results concerns the near transfer observed. In the present study the isomorphic tests showed resemblances, which could easily be recognized by the testees. This is not the case for all isomorphic tests studied in the literature. For instance, Simon and Hayes (1976) used problem isomorphs of the Tower of Hanoi Puzzle, in which the logical equivalence of the isomorphs was much more difficult to see. It is fairly obvious that near transfer is less likely to occur at such isomorphs. Thus it seems that the finding of near transfer is, among other things, a function of the resemblance between problem isomorphs.

A final reservation concerns the finding of a rather high degree of method transfer in the Zambian group. This kind of transfer is the result of a process in which individuals learn general features of the testing situation, e.g., to answer questions with a multiple choice response format, to understand (highly abstract) test instructions, and to work under time pressure. Individuals in Western societies can be assumed to have mastered these skills in the course of their formal education. In non-Western groups with a large test naivety, this complex set of test taking skills will be learned during the tests. The less experience the subjects have with testing, the higher the influence of method transfer can become. It could even be argued that the data analyses used here, in which the data of the three cultural groups are treated in a combined way, can lead to erroneous results as the scores of, e.g., Dutch and Zambian subjects have a quite distinct meaning; in the latter group of subjects the test performance is a function of the intended psychological construct as well as of the ability to deal with tests. Although the score increase from the first to the second administration was considerable, there is not much reason to assume that the Zambian pupils reached the level of overlearning in this ability which is characteristic for Western subjects.

This would have interesting consequences for repeated administrations of tests such as used here. When our view is correct, this would mean that Western groups sooner reach the upper limit of their performance level than non-Western groups. Moreover, score increases in Western groups can be interpreted more easily as they are related to the psychological construct under study, while score increases in non-Western groups reflect the impact of additional learning processes. In order to get a more detailed picture of the latter processes, it may be advisable in future studies to include a larger number of posttests.

In conclusion, it has been demonstrated in the present study that given an appropriate test design different kinds of transfer, ranging from the learning of highly specific test particulars to broad trait-like learning outcomes, can be investigated in a single data set. In the present case the test design involved three tests designed on the basis of explicit item generating rules, while two of these tests were isomorphic. Obviously, this is not the only design in which different kinds of transfer can be studied. However, it highlights the importance of the composition of the test battery in transfer studies. Another consideration in the test design should be what kind of transfer could be expected in a training study. Training studies in which little transfer is expected should use more isomorphic-like instruments, while long extensive training procedures should also include instruments which share less with the primary target of training.

It seems that the general question posed in the introduction as to what is being learned in a training of formal thinking cannot be answered in general. Rather, the pattern of transfer to be found after such a training will depend on, among other things, the duration and effectiveness of the training (cf. Kuhn and Angelev 1976), the age of the testees (Holzman et al. 1976), the degree of test-wiseness of the participants, the general level as well as the dispersion of difficulty of the items of the pre- and posttest, the psychological distance between the target of the training and the test, the absence or presence of common method variance of the training task and the dependent variable, etc. The learning outcome of a training study is modulated by all these and undoubtedly quite a few more factors.

References

Bedderman, T.A., 1973. The effects of training on the development of the ability to control variables. Journal of Research in Science Teaching 10, 189–200.

Borg, I., 1981. Multidimensional data representations: when and why. Ann Arbor, MI: Mathesis Press.

Cantor, D. (ed.), 1985. Facet theory: approaches to social research. New York: Springer.

Chiappetta, E.L., 1976. A review of Piagetian studies relevant to science instruction at the secondary and college level. Science Education 60, 253–261.

Dixon, W.J., 1981. BMDP statistical software. Berkeley, CA: University of California Press.

Ekstrom, R.B., J.W. French and H.H. Harman, 1976. Kit of factor-referenced tests. Princeton, NJ: Educational Testing Service.

Foa, U.G., 1965. New developments in facet design and analysis. Psychological Review 72, 262–274.

Gick, M.L. and K.J. Holyoak, 1983. Schema induction and analogical transfer. Cognitive Psychology 15, 1–38.

Greenbowe, T., J.D. Herron, C. Lucas, S. Nurrenbern, J.R. Staver and C.R. Ward, 1981. Teaching preadolescents to act as scientists: replication and extension of an earlier study. Journal of Educational Psychology 73, 705–711.

Inhelder, B. and J. Piaget, 1958. The growth of logical thinking from childhood to adolescence. New York: Basic Books.

Holyoak, K.J., 1985. 'The pragmatics of analogical transfer'. In: G.H. Bower (ed.), The psychology of learning and motivation. Orlando: Academic Press.

Holzman, T.G., R. Glaser, and J.W. Pellegrino, 1976. Process training derived from a computer simulation theory. Memory & Cognition 4, 349–356.

Kotovsky, K. and H.A. Simon, 1973. Empirical tests of a theory of human acquisition of concepts for sequential patterns. Cognitive Psychology 4, 399–424.

Kotovsky, K., J.R. Hayes and H.A. Simon, 1985. Why are some problems hard? Evidence from Tower of Hanoi. Cognitive Psychology 17, 248–294.

Kuhn, D. and J. Angelev, 1976. An experimental study of the development of formal operational thought. Child Development 47, 697–706.

Linn, M.C. and H.D. Thier, 1975. The effect of experimental science on the development of logical thinking in children. Journal of Research in Science Teaching 13, 49–62.

Neimark, E.D., 1975.' Intellectual development during adolescence'. In: F.D. Horowitz (ed.), Review of child development research, Vol. 4. Chicago, IL: University of Chicago Press.

Poortinga, Y.H., 1975. Limitations on intercultural comparison of psychological data. Nederlands Tijdschrift voor de Psychologie 30, 23–39.

Rogoff, B., 1981. 'Schooling and the development of cognitive skills'. In: H.C. Triandis and A. Heron (eds.), Handbook of cross-cultural psychology, Vol. 4. Boston, MA: Allyn and Bacon.

Royer, J.R., 1979. Theories of the transfer of learning. Educational Psychologist 14, 53–69.

Sarnacki, R.E., 1979. An examination of test-wiseness in the cognitive test domain. Review of Educational Research 49, 252–279.

Siegler, R.S. and M. Atlas, 1976. Acquisition of formal scientific reasoning by 10- and 13-year-olds: detecting interactive patterns in data. Journal of Educational Psychology 68, 360–370.

Siegler, R.S. and R.M. Liebert, 1975. Acquisition of formal reasoning by 10- and 13-year-olds: designing a factorial experiment. Developmental Psychology, 401–402.

Simon, H.A. and J.R. Hayes, 1976. The understanding process. Cognitive Psychology 8, 165–190.

Simon, H.A. and K. Kotovsky, 1963. Human acquisition of concepts for sequential patterns. Psychological Review 70, 534–546.

Spada, H. and B. McGaw, 1985. 'The assessment of learning effects with linear logistic test models'. In: S.E. Embretson (ed.), Test design. New York: Academic Press.

Van de Vijver, F.J.R., in press. Group differences in structured tests. Paper presented at the Advanced Study Institute, Athens.

Van de Vijver, F.J.R. and Y.H. Poortinga, 1982. Cross-cultural generalizability and universality. Journal of Cross-Cultural Psychology 13, 387–408.

Van de Vijver, F.J.R. and Y.H. Poortinga, in press. 'Testing across cultures'. In: R.K. Hambleton and J. Zaal (eds.), Handbook of testing. Amsterdam: North-Holland.

Wing, H., 1980. Practice effects with traditional mental test items. Applied Psychological Measurement 4, 141–155.

Afin de rechercher ce qui est appris dans une étude d'apprentissage sur le raisonnement inductif, trois tâches étaient composées, utilisant deux modalités de stimuli différentes, à savoir des caractères et des figures. Chaque tâche était composée sur base de règles définies explicitement pour la génération de stimuli. Les deux tâches figurales étaient des tâches isomorphes, c'est-à-dire qu'elles ont des structures de problème identiques mais qu'elles utilisent des modalités de stimuli différentes, des figures géométriques simples et des figures de fruits. Ces tâches ont été données aux élèves de sixième d'école primaire aux Pays-Bas, au Surinam et en Zambie. Ensuite, la moitié ne recevait aucun entraînement. Pendant l'entraînement les élèves étaient amenés à résoudre des problèmes en essayant de découvrir la règle pour la génération de stimuli. Quatre résultats possibles de transfert étaient postulés, qui différaient en généralité de transfert des résultats de l'apprentissage. Au moyen d'une analyse multivariée de variance avec mesures répétées, il était possible de différencier les quatre types de transfert: une configuration d'un transfert assez limité était démontrée dans chacune des cultures. En Zambie, une augmentation remarquable des cotes a été trouvée dans la tâche des caractères dans la condition expérimentale ainsi que dans la condition de contrôle. On suppose que cette augmentation a été causée par l'amélioration des aptitudes pour ce genre de tâche, acquise pendant l'administration initiale des tâches.